48

OPERATIONS RESEARCH

Norman C. Dalkey

October 1967

P-3705

3

THE RAND CORPORATION 1700 MAIN STREET, SANTA MONICA, CALIFORNIA 90406

11-30-67

ERRATUM

P-3705, Operations Research, by Norman C. Dalkey, October 1967.

Page 9, Figure 1. Capacity on arc D F should be 1.

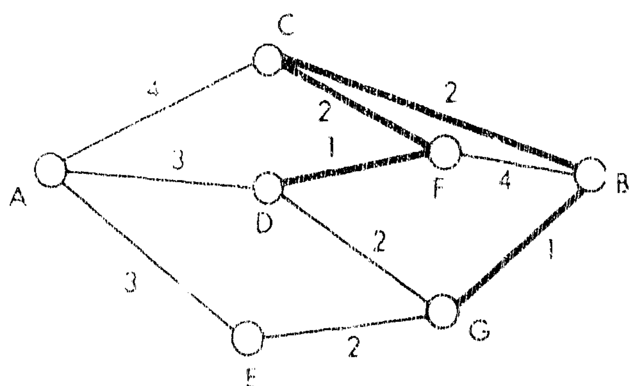

Fig. 1

The Reports Department

OPERATIONS RESEARCH

Norman C. Dalkey*

The RAND Corporation, Santa Monica, California

The notion has been around at least since the time of the Greeks that the world can be described and dealt with rationally, that it can be managed with numbers and logic. This notion has had its ups and downs in the history of western thought. But only a few scientists and philosophers made much of the idea at any time, until recently. Men of affairs presumed that numbers played a limited role in their concerns. Numbers were valuable for accounting and for measuring; but most of the world of affairs was a world of doing, of cut-and-trying, of judgment and good sense. Science could offer some very useful items and processes that could be turned into new products and manufacturing methods. But it was the vision of the entrepreneur that stoked the machines of progress, the military insight of the commander that won the day; and the wisdom of the politician that kept the ship of state off the rocks.

Operations analysis is a challenge to this point of view. It suggests that a very wide area of human affairs

---

*Any views expressed in this paper are those of the author. They should not be interpreted as reflecting the views of The RAND Corporation or the official opinion or policy of any of its governmental or private research sponsors. Papers are reproduced by The RAND Corporation as a courtesy to members of its staff.

This paper was prepared as a chapter for an anthology, "The Uses and the Spirit of the Mathematical Sciences," sponsored by The Conference Board of Mathematical Sciences.

is subject to rational analysis and control. In this respect, it more nearly encapsulates the spirit of the post-war world than any other formal discipline. It is intensely pragmatic, concerned with direct operational decisions in industry and government. But also, it represents the attempt to extend standards of rigor, objectivity, and conceptual generality to subjects which have been the domain of judgment and raw experience. Perhaps the sharpest difference between operations research and traditional science is that operations research is frankly prescriptive. The traditional role of science has been to describe the world, as accurately and completely as possible. The operations analyst is primarily concerned with the search for better ways to manipulate the world. Accurate description is, of course, a prerequisite for finding improved courses of action. But it is not sufficient. In the language of the Greeks, operations analysis is concerned not only with the true, but also the good. It offers calculated planning in the place of informed muddling through.

Operations research is a very young discipline. It obtained its status and a name during World War II, when a remarkable assemblage of individuals ranging in background from lawyers and sociologists to mathematicians and astronomers put their intellectual skills to work on military problems. Their interests ran the gamut of modern warfare--improved accuracy in strategic bombing, preferred tactics in tank battles, better ways to search for submarines with aircraft and surface ships, more effective anti-aircraft defenses... The general aim of the studies was not new; commanders have been interested in improved methods of conducting military operations since the dawn of warfare. What was new was the approach. Despite the diversity in background of the analysts a common pattern emerged. The problems were approached with the traditional tools of

science--careful observation, theory formulation, and where possible, mathematical manipulation of the theory to find better ways of conducting operations.

The impact of this new approach on methods of warfare was rather modest, measured on the scale of the revolutionary effects of physical science and technology. There was nothing comparable to RADAR or jet aircraft or the atomic bomb. But the successes were real; and the demonstration that complex systems of machines and men are amenable to rational analysis opened wider the windows of the scientific outlook on the world. It was quickly recognized following the war that the new approach was applicable to a host of practical problems, not only in the military, but throughout industry, business and more recently to government both national and local. The stream of applications to this diversified subject matter rapidly became a flood. There are now two international professional societies devoted to the promotion of operations research, as well as a society specialized to military analysis. The activity has spawned a number of related disciplines--systems analysis, cost-effectiveness analysis, management research--which are sometimes difficult to distinguish from the parent except in name.

Like most new intellectual disciplines, operations research has expanded without much regard to boundaries. It has ebulliently spilled over into areas related to economics, engineering, and social policy. The conceptual structure has remained loose, corresponding more to an attitude and a way of attacking a problem than to a theory. Practitioners utilize whatever formal areas of mathematics seem useful, and are happy to apply theoretical structure from other sciences when an analogy can be found. What has been most exhilarating about the development of operations research has been the continuing proof that the net of science can catch a multitude of bizarre subject matters...

steel mills and banks, cases in courts of law and messages in a telephone system, traffic on freeways and urban growth. It is possible to describe these diverse processes in precise, compact, and sometimes elegant fashion; to compute the behavior of machines, materials and men in swift summary or lavish detail (if you have a high-speed computer!) and to manipulate that behavior in conceptual experiments to generate new policies and procedures.

MODELS AND ENTERPRISES

The subject matter of an operations analysis is usually an enterprise--an industrial process, a military campaign, a transportation system. The enterprise will be characterized by a set of operations involving machines, materials, and men. It has a goal or goals determining the value of the output, and a policy, a set of rules for conducting the enterprise. The heart of an operations analysis is the creation of a model, a precise manageable description of the processes involved. The hooker here is not so much the precise as the manageable requirement. In general this means two things--numerical and succinct. To be succinct, the description must abstract from a host of features of the enterprise which are irrelevant to the problem being posed. It often requires a great deal of ingenuity to "find" such a description.

On the other hand, the creation of a useful model in operations research does not require the kind of profound digging that is needed to ferret out the role of genes in animal heredity, for example. A major reason for this is that most enterprises are artifacts. Or at least, the kinds of enterprises that operations research has found tractable are the sort that have been designed. A savage, wending his way through a forest, may trace out a path that defies simple description. But a highway is constructed with careful forethought, and is likely to have geometrical simplicity. In the forest traversed by our savage, a tree growing under the influences of wind, sunshine, soil and moisture distribution, and the competition of other plants, will take on a complexity that only very subtle analysis like that of genetics can rationalize. But a telephone system, growing under the rising demand for communications, is structured according to the rules of communication engineering. The laws describing the nature of these systems are only in part the laws of nature. Many are the laws of design.

This feature does not relieve the analyst of the need to begin with a careful scrutiny of the behavior of the enterprise. It is a rare enterprise that operates precisely as intended. It is necessary to watch the processes in action, to collect numerical data, and base the model on the observations. Frequently, the process of carefully observing an enterprise will produce surprises.

A delightful case in point is the experience of the Danish analyst Arne Jensen (1). He was asked to assess the increased risk of collision resulting from mounting boat traffic in a narrow channel between Denmark and Sweden. The channel carried both ocean-going and local traffic, in particular ferry boats. The density of traffic threatened to double within a few years. It might be worth building a bridge to replace the ferries. But how to estimate the increased risk of collision due to a doubling of traffic? Collisions are sufficiently rare so that counting-type statistics give little information.

Jensen and his colleagues took time-lapse pictures of the scope of a surveillance radar that monitored the traffic in the channel. As a movie, these pictures produced a speed-up of two hundred and fifty times. Study of the movie was unrewarding, and even a panel of experienced seamen could make little of it. But Jensen noticed that at various times during the showing there were moments when tension appeared in his audience, straining forward in seats. When the incidents producing this audience reaction were examined they did not turn out to be near misses, nor intricate tangles of boats; they were mostly cases where three boats were involved. The "rules of the road" for boats are written to deal with the situation where two boats are on potential collision courses. When three boats become involved, the rules are no longer unambiguious.

Jensen had his criterion; the increased hazard could

be measured by determining the increased number of cases where three boats could become involved in an avoidance problem. A simple calculation showed that a two-fold increase in traffic would produce an eight-fold increase in the number of three-boat incidents. Suddenly the bridge looked very good indeed.

Observation has to go hand in hand with abstraction--the selection of what to observe. To a visitor, a steel mill is a vast assemblage of thunderous activities. To an operations analyst it is a great deal simpler. The glowing blast furnace where iron ore is turned into iron becomes a simple input-output table: x tons of ore, y tons of lime-stone (a fluxing material), z tons of coke (a fuel) and t hours of time, produces w tons of molten iron and v tons of slag. The other components of the mill can also be represented in input-output form. The great rolling machines can be expressed in the simple relationship: one bloom (a large, red-hot block of steel), k kilowatts of electrical energy, p passes through the mill each taking t minutes, one skilled operator, produces one billet (a longer, flatter, more manageable shape). Of course, for some studies more detailed descriptions will be desirable, but the drama, the earth-shaking noise, the spectacle of the glowing megablock of steel undergoing transformation, can safely be ignored. Stringing the input-output tables for all the components together produces a larger table: raw materials, labor, and energy as inputs, finished shapes after a time lag. With such tables many crucial features of operating the mill can be computed, such as preferred mixtures in the furnaces, the requirements in materials and labor to fulfil a large order for structural shapes, the most efficient dove-tailing of activities to simultaneously fill several orders.

The input-output model has turned out to be applicable to very many enterprises. It is also a useful way to describe

entire industries, and even the total economy of the nation.

For other enterprises different forms of model are more appropriate. One of the continuing pleasures in operations research--as in other sciences--is the way that models developed to describe one kind of phenomenon turn out to be useful to describe a very different kind of activity. An example is the use of hydrodynamic equations to describe traffic flow. In many respects the movement of traffic on a highway resembles the flow of a liquid through a channel. The theory of fluid flow is well known, the theory of traffic flow just beginning. In one satisfying application of the theory to traffic flow in the tunnels that connect Manhattan Island to New Jersey (2) it was noted that slow-downs behaved very much like standing waves in a channel. One car slowing down would cause the following to slow down and the disturbance would persist long after the original laggard had left the tunnel. The analogy suggested that perhaps some method of segmenting the traffic would break up the disturbance. It was not difficult to compute the traffic density at which waves are likely to occur, and a procedure was evolved so that when the traffic reached this critical density it was periodically interrupted for a few seconds. Thus the traffic moved through the tunnel in a sequence of bunches. Any wave which formed was limited to a single cluster, and died out as the cluster moved on. With this procedure in effect, traffic flow in peak hours increased over 10%. In addition a number of unpredicted (but pleasant!) side effects resulted: ventilation requirements went down due to the decreased slowdown-acceleration cycles, collisions declined, traffic at the tunnel entrance was less congested.

For many enterprises, a model based upon the analysis of waiting lines (called queues in England, hence the name "queuing theory") has been illuminating. Although the prototype is humble, the ramifications of the theory can

lead to some of the roughest mathematical problems. The theory is primarily statistical. Additions to the waiting line arrive at random times. Their needs will vary so that time at the service station will also be random. A number of concerns can arise. Excessive waiting lines can arise. If additional service units are added to reduce the queues, costs can mount, and the units will frequently be idle. Balancing these concerns means walking a statistical tightrope. Enterprises ranging from communications networks to air traffic at airports have been analysed with this type of model.

There is almost always more than one way to trap an enterprise in a model. Which trap you will use depends on what you intend to do with the catch. One form of model (which just happens to have among its many aliases the name "net") applies to many of the processes which can be analysed by queueing theory, but gives a different type of answer. It is usually called graph theory. A graph, in the abstract, is a collection of two kinds of things, nodes, and a set of arcs connecting some pairs of nodes. Represented on a sheet of paper, a graph consists of a set of points and lines between some pairs of points, as in Fig. 1.

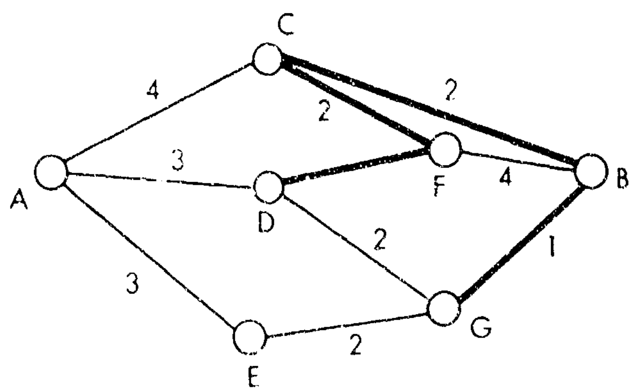

Fig. 1

For the operations analyst, the lines may represent communication links, transportation routes, assembly-line conveyors, or more abstract things like the occurrence of a significant advance in a research and development project. The points can be switching centers, way stations, machine stands, or stages in a development process. For most uses the lines will have numbers attached to them representing capacities, distances or lag times. The value of graphs as models of enterprises derives in large part from a basic theorem due to Ford and Fulkerson (3). Suppose you wish to find the maximum amount that can be shipped from point A to point B, taking into account the capacity limitations, but allowing any combination of routes. A cut between A and B is a set of lines which, if they are removed, interrupt all routes between A and B. One cut would be all of the lines coming out of A. Another would be all lines. Add up the capacities on all the lines in a cut and call that the capacity of the cut. The theorem states that the maximum amount that can be shipped from A to B is equal to the capacity of the cut that has the minimum sum. In Fig. 1 this cut is indicated by the heavy lines.

For a small network like the one in Fig. 1, finding the minimal cut is no great matter. But for large networks, the number of possible cuts can become enormous, and requires special methods that are the subject of another section.

To sum up this section, a model is a form of description--compact, shorn of irrelevancies, precise. But a description is not the final goal of operations analysis. The solution of operational problems is what is wanted. To discuss this, we need a slight excursion on the subject of the uses of mathematics.

BETTER, BEST

Mathematics has often been characterized as the language of science, a way to describe events in simple, quantitative terms. But it is much more. Newton's law of gravity, $F = GmM/r^2$, is a lucid way of saying that the force F attracting two bodies of masses m and M respectively, is directly proportional to the product of the masses, and inversely proportional to the square of the distance r between the two bodies. The constant G is a way of keeping the units in balance on each side of the equation. The law describes the gravitational attraction between any two objects--Mars and Jupiter, an orbiting satellite and the earth, a comet and the sun. It allows assuming for all practical purposes that the gravitational attraction between two atoms is negligible.

But if that were all we could do with the law of gravity, it would be of limited interest. It is the fact that from such a simple, unassuming statement, we can derive (with the help of a few more equally simple laws) consequences of a remarkably intricate sort that makes the laws so profound and powerful. Out of the laws, the entire path of a spaceship looping around Mars, or the periods of the planets in their measured wheeling around the sun, can be evolved. We can derive such fascinating facts as: if the earth were a hollow shell, any object inside the shell would have no weight at all. These, and a myriad other complexities are contained in the simple basic laws. The device which unlocks the storehouse of consequences is mathematics; it is the loom that weaves the few, elementary threads into endless and intricate patterns.

The pattern that the operational analyst prizes above all is contained in the words "optimum" or "maximum." And well they might prize it, for optimization problems are among the most difficult in the entire field of mathematics. As long as the enterprise is very simple and can be described

as if all the quantities involved were indefinitely divisible, a large body of mathematical techniques is available. (The "as if" is significant here. The quantities don't have to be fragmentable as long as no harm results from saying they are.) Consider the case of a firm that manufactures a single product. Let's say it can produce up to 100 units a day with available facilities, but any additional units would require overtime, and beyond 125, additional facilities would have to be rented. The return (or profit) for various amounts produced might look like Fig. 2. (Profits are negative below 60 because of "overhead.")

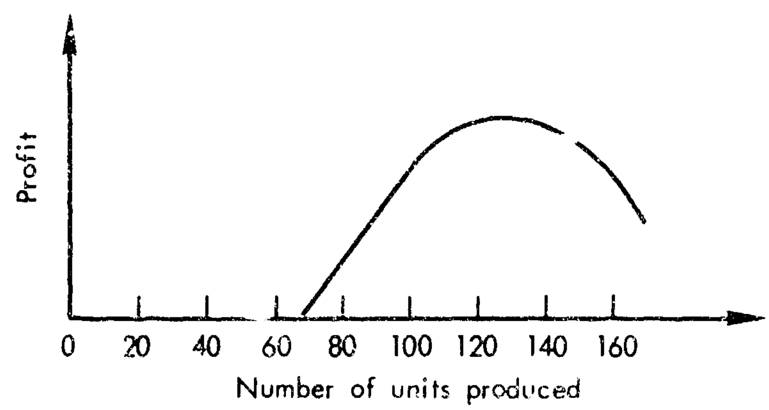

Fig. 2

It is easy for the eye to pick off the point where the firm would receive its maximum return from this graph. It is also easy, if the relationships are "nice," to pick out the maximum point by formula. A comparable situation for two products might look like Fig. 3.

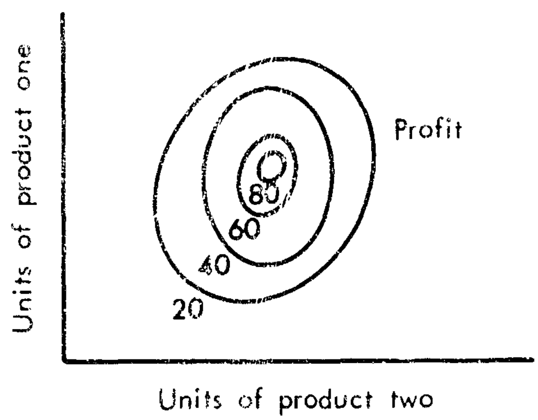

Fig. 3

Here the ovals are the contour lines for a profit "hill" rising out of the page. Again, the human eye can find the peak with no trouble, and techniques exist to search for it with formulae. This game ends abruptly at two, as far as the human eye is concerned, but can be continued by analytic means for many more dimensions. Real trouble begins when special constraints are imposed. For example, suppose our company knew that its share of the market would allow it to sell only A units of product one, and B units of product two. The situation might be as in Fig. 4. The peak is now irrelevant. Methods of "hill-climbing" for multi-dimensional hills have been devised for high-speed electronic computers that can traverse the conceptual slopes at something approaching the speed of light, but they slow down when they are not zipping for the peak and must patiently explore the side ridges for a constrained maximum.

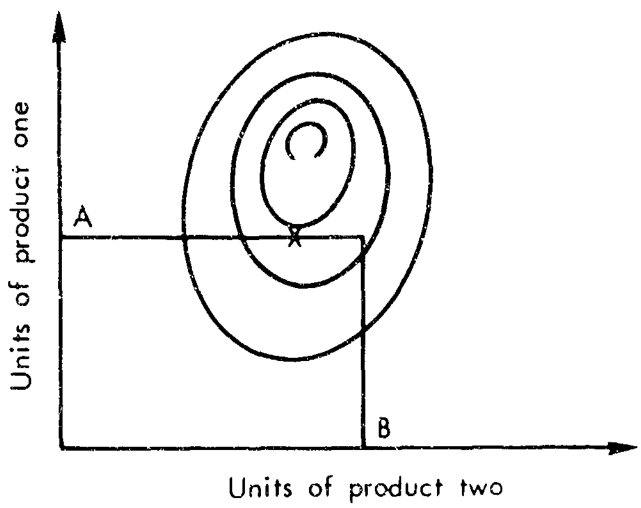

Fig. 4

The problems become really severe when it is not possible to make the simplifying assumption that the quantities are infinitely divisible. In the case of the graph theory model mentioned in the previous section, the number of routes between two points is an integer and there is no way to "smear" them into a continuous quantity. In some very general sense, the occurrence of integers simplifies the situation, but in practice the trouble is that the number of cases gets out of hand very rapidly. The situation is illustrated by a famous (or perhaps "notorious" is a better description) problem known as the Travelling Salesman Problem. A traveling salesman wants to make a round trip visit to a number of cities. He needs to pass through each city only once. In what order should he visit the cities to make his total trip distance least? If there are only a few cities, he can try all possible tours and choose the one that is shortest. But a little arithmetic quickly discourages this

approach if the number of cities is at all interesting. For four cities the number of possible tours is three, for five cities, twelve; for six cities the number is already sixty. If the salesman wanted to visit the capitols of all the fifty states, he would have to scan a list of 3×10^{55} tours (3 followed by 55 zeros!). I'm told that traveling salesmen have much more interesting ways to spend their time. The trouble is there is no completely trustworthy way to weed out the longer tours by looking at parts of the trip.

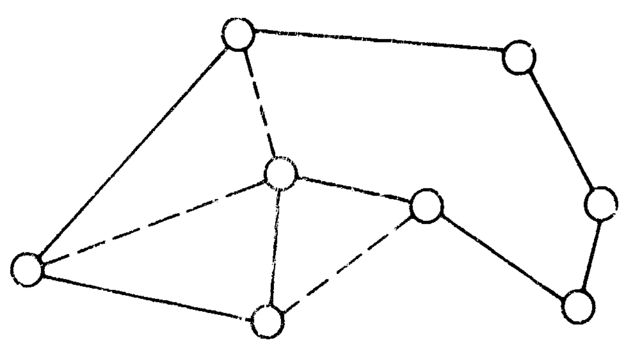

Fig. 5

Fig. 5 shows one tour through eight points. It is not an optimal tour. For example, the changes introduced by the dotted lines would shorten it. The reader might be interested in trying to shorten it further.

To see this, we can look at a much more compliant problem, but one that at first sight appears to be very similar to

the problem of the traveling salesman. Suppose our traveler only wants to get from some city A to another B, and wants to choose the shortest route. This problem turns out to be very manageable by a technique called dynamic programming.

Consider the map in Fig. 6. There are eight ways to go from A to B (without doubling back). We could list all eight and select the shortest but I have already hinted that that can't be the right way. Another attack is to look at what is called a policy. In this case a policy consists of deciding for each city how you are going to leave that city. A-C, C-F, D-G, F-H, is such a policy. (We don't have to decide for E, G, H, I, or, of course B.) A policy determines one and only one route from A to B. Now it so happens that there are sixteen policies (two possibilities at each city, four cities). It looks like

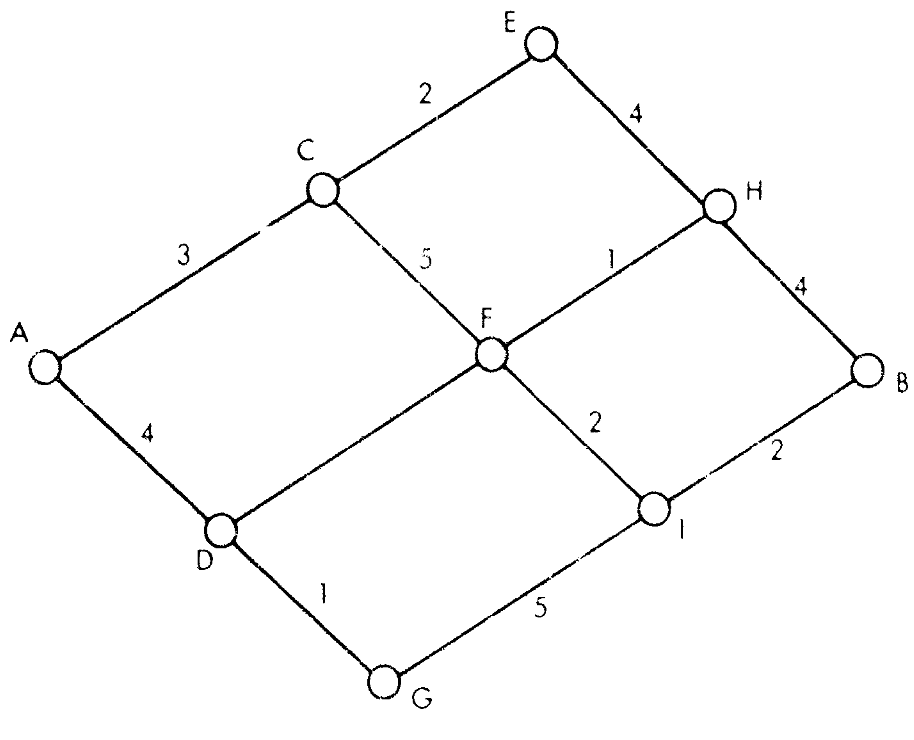

Fig. 6

we have leaped from the frying pan into the fire. But the nice feature of a policy is that we can change it piecemeal. For example, the distance from F to B is 5 via H and 4 via I. We change the policy at F to I. At C the distance is 9 via F and 10 via E. We leave C alone. Continuing in this fashion we can sucessively improve the subpolicies until the problem is solved.

For very many problems, it is possible to proceed in the fashion of our simple example--that is it is possible to define a policy which consists of a number of subpolicies, and the problem can be solved by successively improving the subpolicies. The difficulty with the traveling salesman problem is that, although we could define policies and subpolicies for it, the solution cannot be guaranteed by dealing with subpolicies alone.

Very many problems can be dealt with by the methods of programming. Where input-output models of the sort described for the steel mill are appropriate, programming techniques are usually what is needed to solve scheduling or costing problems (4), (5).

GOOD

Discussing optimization, I took it for granted that what the analyst is trying to maximize is clear. This is rarely the case in operations research problems. For enterprises where the aim is primarily monetary profit, the determination of a value scale is relatively easy. There may still be major problems of measurement. This is particularly true if the analysis is concerned with a future operation--e.g., a firm is trying to decide whether to include a new product in its offerings to the public. A study for the Air Force discovered that underestimates of the cost of future weapons systems of the order of three to ten times were not unusual.

Consider a municipal bus service. Clearly money plays a big role in its life. If an analyst can find a way to produce the same service at reduced cost, he is in for a medal. But suppose the company is thinking of changing its service (perhaps under the pressure of declining use--a rather common fate for public transportation these days). What constitutes improved service? Unfortunately, many things. Speeding up individual bus trips from station to station would be good, increasing the comfort of buses would be nice, perhaps more routes so that bus stops could be closer to origins and destinations, reduced fare?--and so it goes. One study to remain nameless pointed out quite seriously that at least the travel time from origin to destination by taking the bus should be less than the time required to walk the distance. Trading off these competing goods is like navigating in a fog with a restless compass.

This situation is not rare. Consider the plight of the operations analyst who has been requested to evaluate the Post-Apollo program of NASA, (After the Moon, what?). Leaving aside all of the many uncertainties and the vast complexity of programs that can be devised, how is one to compare, say, landing a team of scientists on Mars

with sending a fleet of unmanned space probes to explore the outer planets and their satellites? A hospital has to decide whether to buy an intensive care unit for cardiac patients or a computer to keep tabs on the shifting supply of drugs at nurses' stations.

There are a number of tricks to the trade that soften the harshness of some of these value conflicts. One is a simple precept that can relieve a vast load of guilt. It can be summed up in the slogan "suboptimize!". No matter how tantalizing the dream may be to "squeeze the universe into a ball and roll it toward some overwhelming question," an analyst cannot wrap up everything in a cosmic systems analysis and forever after follow The Policy. Since all that can be done is suboptimize anyway, it makes sense to tackle problems that can be posed with relatively clear objectives.

Sometimes life is kind, even to the analyst. Faced with competing criteria, alternatives just might be found that turn out best on all criteria. That isn't the kind of good luck that you base the fate of a study on, but it happens. This is known as the principle of dominance. When it doesn't work in a positive fashion, it is often useful in a negative way. If one alternative X is better than another Y on all criteria, you can certainly forget Y.

Values appear to have an inherent reluctance to being measured. Again, with luck you can find a stand in--something that is measurable, and at least varies in the same way as what you really are interested in. The United States maintains vast nuclear forces to deter potential enemies from attack. There is no measuring rod for deterrence. A standard approach is to compute the number of enemy deaths our forces could inflict if the enemy attacked first. Potential enemy deaths is not the same thing as deterrence. Hopefully they are monotonically related, i.e., deterrence goes up when potential deaths increase.

When the apples and oranges situation is unresolvable, all is by no means lost. A case in point is the now justly famous Cost-Effectiveness type of analysis pursued by the Department of Defense. Military effectiveness is not directly comparable with dollars, and alas, you cannot simultaneously maximize effectiveness and minimize costs. If the budget process sets a level beyond which you cannot spend, then you can fix your attention on effectiveness, and maximize it for the given budget. Conversely, if the enemy, by his force posture, makes a certain level of effectiveness imperative, then you can seek a posture that will achieve that level of effectiveness at least cost. Even if these two modes are not open, you can still plot effectiveness against cost and nine times out of ten the curve will look like Fig. 7--the more you spend, the smaller return for an additional dollar, and somewhere in the shaded area you will say, "It's just not worth putting more money into this one."

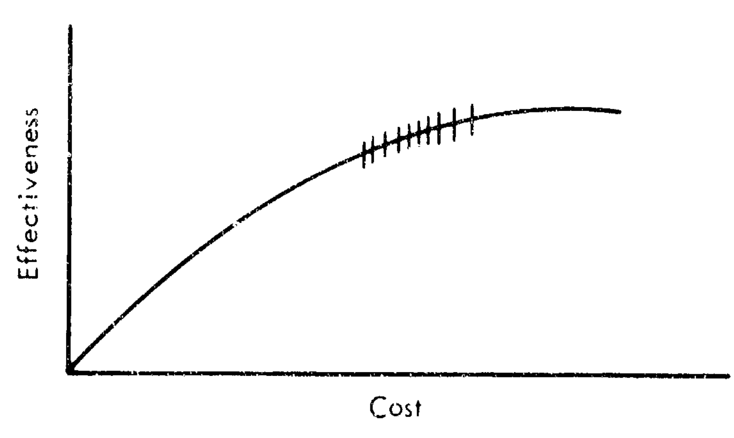

Fig. 7

A potentially powerful tool for measuring values is afforded by utility theory. This is a theory that grew out of careful contemplation of gambling behavior. Many profundities have been uncovered by the study of gambling and related activities: statistics, the theory of games, psychological learning theories, a good deal of economics. The reason is that gambling behavior involves decision making in risky situations which is similar to decision making in many human enterprises, but is much easier to analyze. With regard to measuring values, gambling enters in the following way. In a risky situation, some outcomes can be expected only with a certain probability. If the decisionmaker knows what he prefers among these uncertain outcomes, then the probabilities themselves give a numerical scale. This is true whether the outcomes themselves are numerical or not. To take a very simple (and not very realistic) example, a young man is contemplating a choice between three lovelies of his acquaintance, Joyce, Claire, and Mary. As matters stand he runs a risk in concentrating on any one. He can perform the following "thought-experiment." Suppose he could enter a lottery where he had a choice be between Clair with certainty, and a fifty-fifty chance of Joyce or Mary, that is, if he chooses the lottery ticket he has an equal chance of winding up with Joyce or with Mary. Which would he prefer? If he would prefer Claire to the lottery ticket, this presumably means he rates Claire closer to Joyce than to Mary. If he prefers the lottery ticket, then Claire is closer to Mary. Refinements in the imagined lottery ticket can lead to pinning the numerical comparisons down even closer.

If the preferences of our subject are consistent then the probabilities are a measuring stick for "desirability." Consistent means mainly that the preferences do not go around in a circle; he doesn't prefer A to B, B to C, C to D and then D to A.

The theory of utility has clarified a number of puzzles in human behavior. For example, most lotteries are "unfair" in the sense that the cost of the ticket is greater than the expected return. Why do people buy lottery tickets? A reasonable answer is that the <u>value</u> of a large sum of money to the player is greater than the numerical amount of the prize.

In practice, the theory of utility has been more useful as a conceptual guide to the analyst to help him think through his problem than it has been as a value-measuring device. The reason is that present methods of obtaining judgments of preference over very many different possible alternatives are very cumbersome. But Rome wasn't built in a day.

BETTER FOR WHOM?

Up to now I have been discussing enterprises as if they existed alone like Robinson Crusoe on his island, and the only problem was to find what was good for the enterprise and devise ways of increasing that good. The conceptual problem here has a splendid simplicity. But most enterprises do not exist in isolation; there are competing firms that are eager to increase their share of a market; the enemy has as much to say about how a military campaign will go as we do; the members of a legislature represent constituencies with different interests. The fate of an enterprise will depend upon the decisions and actions of other enterprises. In deciding what to do, Enterprise A must take into account the policies of Enterprise B, which must take into account the policies of A ... and we have what is known as a rat-race.

Help in this vertiginous situation has come from the Theory of Games (6) (7) a very profound theory of what might seem at first sight like a superficial subject. But social games such as bridge, chess, and poker are very good objects to study because they involve in a clear way the elements of competition and mutual adjustments of strategy that are present, but more obscure, in the serious games of business and war.

The help consists in two basic insights--a precise notion of *strategy*, and a precise notion of "mutual taking into account." A strategy is, first of all, a plan, but equally important, it is a plan that furnishes the player with a response, *for every possible course of action open to the other players*. To be a strategy, the plan need not be a good one, just that it be complete. This might appear to be the reverse of brilliant, in fact downright pedestrian. However, the importance of the notion arises from the insight it gives into the structure of a game. In the game of tic-tac-toe, a strategy for the first player might start out:

I will put my first cross in a corner. If the second player puts his zero in the center, I will put my second cross in the opposite corner. If he plays anywhere else, I will put my cross in the center. Then if .... A little computation shows that there are something like 10^{127} such strategies for the first player! This astronomical number--it is not far from the English Astronomer Eddington's estimate of the number of elementary particles in the universe--is actually tiny compared with the number of strategies in the game of chess. Some notion of the power of the human mind is contained in the fact that most elementary school children have a fairly good idea of how to play tic-tac-toe reasonably.

Since the publisher would object if I wrote down all strategies for tic-tac-toe, it is politic to use a much simpler game as an illustration. Suppose two players A and B each write down a number between one and three. They simultaneously disclose their numbers, and Player A's score is the difference between the two, except when A has written "one" and 3 has written "one" or "two," in which case the score is 3 and -3 respectively. This sounds a little complicated, but it can be very neatly represented by the array in Table I.

Table I

A \ B	[1]	[2]	[3]
(1)	3	-3	-2
(2)	1	0	-1
(3)	2	1	0

The score for player B is just the negative of the score for Player A.

An array like Table I is called a payoff matrix. In general a payoff matrix for a game consists of listing all the strategies of one player as rows, all strategies of the other player as columns, and for each pair of strategies giving the score (winnings, payoff) of one of the players. In the simplest games, the payoff to the other player will be just the negative of the payoff to the first, in which case the game is called zero-sum. This is one way to precisely describe a game.

How should the two players play in the game of Table I? Player A might try writing down (1) in hopes of getting the 3. But then Player B would like to write [2], getting 3 for himself. But then Player A would rather play (3), and so it goes... Nevertheless, there is a reasonable way for both players to play in this game. Player A can reason thusly: If I play (1) Player B could play [2], and I would lose three points. If I play (2), Player B could play [1], and I would lose one point. If I play (3), Player B could play [3], and I would lose zero points. The safest strategy to play is (3), because I will get at least zero and perhaps more. Player B, using the same line of reasoning, would find [3] his safest strategy. The situation is that Player A can guarantee himself at least zero, and Player B can guarantee that he will not get more than zero. It is reasonable to assume that this is the best either can do.

Taking account of the ther player's taking account is thus accomplished. A finds his best strategy by choosing the one that guarantees the most he can get if B is assumed always to respond with the strategy that will do best for B, and conversely B chooses assuming A will do the same. This criterion, called "minimax," is an intuitively satisfying way to select strategies, when the criterion gives

the same result for both players. This is not always the case. Consider a very simple game with only two strategies described by Table II.

Table II

B / A	[1]	[2]
(1)	1	-1
(2)	-1	1

If Player A plays strategy (1) the minimum is -1. Similarly the minimum is -1 for strategy [2], so Player A can guarantee himself only -1. On the other hand, if Player B plays his strategy [1], the maximum is 1, and if he plays strategy [2], the maximum is [1] so Player B can guarantee only that Player A will get 1. The two are not the same.

In this case, in order for Player A to do better than -1, he must make use of a more general type of play called a mixed strategy. If he tosses a coin to decide between strategies (1) and (2), he will get 1 half the time and -1 half the time, no matter what Player B does, and so his average payoff will be zero. Conversely, if Player B tosses a coin, he will get 1 half the time and -1 the other half, and his average will also be zero. With these more general "mixed" strategies, the maxmin is again the minmax and the game has a solution.

The fundamental theorem for finite zero-sum games, states that if mixed strategies are allowed there will always be a solution.

There is general agreement, that the minimax solution of zero-sum, two-person games is a satisfactory resolution

of the problem. That doesn't mean, of course that it is easy to find the good strategies for all games. On the contrary, it is extremely difficult to find solutions to specific games. The major reason for this is, of course, the enormous numbers of strategies for interesting games. No one has attempted to write down even one strategy for chess, and the mere thought of writing down the payoff matrix is appalling. If the whole universe were the blackboard, it just might be possible.

What progress has been made in analyzing games has come from using every trick in the trade to reduce the size of the matrix that must be examined. For some quite simple games, it is possible to approximate the game with continuous quantities, which as we saw earlier, often brings the problem into well-traveled territory. But the number of games that have been solved is still agonizingly small.

The simple change from zero-sum to non-zero-sum games brings a major shift of emphasis. As long as the game is zero-sum, there is no point in cooperation among the players. But as soon as the game is not zero-sum, cooperation becomes strategically reasonable.

Consider the game in Table III, where the first number in the pair gives the payoff to Player A, and the second number to Player B

Table III

A \ B	[1]	[2]
(1)	2,2	0,3
(2)	3,0	1,1

This matrix has become quite famous, under the name "prisoner's dilema."* The dilema arises from the fact that strategy 2 is uniformly better for both. If Player A plays (2) he gets 3 rather than 2 if Player B plays [1], and he gets 1 rather than 0 if Player B plays [2]. Thus, if each player is selfish and plays only according to his own narrow interests, they wind up with 1, 1. But they could have achieved 2, 2, better for both, by cooperating and playing (1) [1].

Thus, in non-zero-sum games, it is possible to improve one's prospects by cooperating with other players, or as it is usually put, by forming coalitions. Strategy now has a new dimension, making agreements. Unfortunately no theory as satisfactory as the minimax solution exists for non-zero-sum games. There are a number of separate solution concepts which appear satisfactory for certain situations, but not for all. A theory of bargaining situations has been developed based on the notion of threat. Other solution concepts make use of the notion of standards of behavior, or social custom. All of these have turned out to be useful in analyzing restricted kinds of social and political situations.

One solution concept that has turned out to have more than academic interest is the Shapley value (8). In general terms, this concept attempts to measure the value to a player of being in a particular game. The evaluation

---

*(The name arises from the playlet: Two prisoners implicated in the same crime are being interrogated separately. Each is offered his freedom if he will testify against the other, and the other doesn't confess. The only evidence available is the statements of the prisoners. If they cooperate, and neither confesses, then they have a good chance of getting off, represented by the 2, 2 at (1) [1]. If one confesses and the other doesn't, the informer gets 3 and the silent one gets 0. If both confess, they get reduced sentences, 1, 1.)

can be made by examining the role of the player in affecting the fortunes of the various coalitions he can ente . If he is not crucial to given coalition, i.e., the coalition could do as well without him, he is not credited with any gain from that coalition. More generally, he may sometimes be crucial, and sometimes not. Roughly speaking, his score is made up of the average amount that he contributes to each coalition he can enter.

For very simple kinds of games where the question whe whether a coalition will win depends solely on its size, for example in voting, the Shapley value can be computed by determining the average number of times a given individual will be pivotal in swinging a vote. This concept has turned out to be of practical concern in the implementation of the Supreme Court's decision for "one man-one vote" in the reapportionment of state legislatures. The Court decision could be most directly implemented by assuring that each member of a legislature represent precisely the same number of voters. In practice this is difficult because it would require "unnatural" districts, e.g., mixing together several communities with different interests. One resolution that has frequently been suggested is proportional representation, keeping the natural districts, but giving the representatives power in proportion to the population of the districts.

This scheme, as it turns out, has certain hazards (9). Depending on the pattern of district sizes, variations in the actual voting power of the representatives (as measured by the Shapley index) can be considerable. In the extreme case, some representatives can literally be dummies--have no real effect on the outcome of voting It is not difficult to adjust the districts so that this hazard is overcome and a fairl uniform degree of power assigned to the legislators. But there is an additional problem. If the Supreme Court's decision is interpreted to apply to the voting power

of the individual, and not to his representative, then even when the voting power of the legislators is uniform, wide variations can exist in the power of the constituents. In general, there is a bias toward the larger districts, and roughly, the power of the individual is proportional to the square root of the population of the district. Thus, if one district is four times as large as another, so that in the proportional scheme, the representative of the larger district would have four votes to the smaller's one, then an individual in the larger district will have about twice the voting power as an individual in the smaller district.

Despite the apparent abstractness of these considerations, there have been several court cases in which they have been tested, and in general the courts appear to look favorably on the Shapley value as an interpretation of the one man-one vote criterion. The same considerations are playing a significant role in congressional hearings concerning reform of the presidential electoral college.

Game theory has extended enormously the conceptual apparatus for analysing economic, social, political, and military situations where conflict, competition, threats, and cooperation play crucial roles. It has been hampered in applications by the excruciating difficulty of solving realistic games, and by the multiplicity of solution concepts in non-zero-sum games. Where solutions have been found, they have carried surprising conviction.

EXPERIMENTS WITHOUT LABORATORIES

The delight of an operations analyst is a problem where good, extensive data about the enterprise exists, where a simple and compliant model can be constructed, and where the aims of the enterprise are crystal clear so that recommendations for new courses of action are as obvious as 1 + 1 = 2. It is sometimes like that--but more often not.

Many things happen on a highway that do not fit the hydrodynamic model discussed earlier. Collisions between cars do not resemble collisions between molecules. Diurnal variations in traffic density are far from the quasi-steady states that make the liquid analogy useful.... These obtrusive factors can be taken into account by a more detailed description of what is happening; by recognizing the differences between different cars and different drivers, by examining the specific geometry of a freeway on-ramp, by including the difference between night and day. These things can be done, but in order to do so it is necessary to give up the neat succinct theory and turn to something called simulation.

To simulate means to build something that is similar to what you are interested in, but something that is easier to study. There are many ways to do this. Aircraft can be simulated by models in a wind tunnel, and ships by models in a test basin. A military campaign can be simulated by a field exercise, a "war game." But the most widely employed method of simulation is by computer.

To teach a high speed electronic computer to pretend that it is a stream of cars along a highway, or a set of law cases being "processed" by the courts, or a succession of telephone conversations passing through a switching central, may seem a little strange, but simulations of this sort make up a large part of operations research. The technique might be called the cinematic approach. As

you know, a moving picture does not really move. It consists of a rapid series of snap-shots which, when projected as a succession of still images on a screen, gives the illusion of motion. For the "scenario" of the highway drama, you furnish the computer with a list of all the cars on the highway, their type, their speed, and their location at a given time (the "opening shot") and a set of rules saying where each car will be at the next instant, depending on its speed, type, kind of driver, and the traffic situation around it. The computer goes through the tedius bookkeeping of changing the description according to the rules, and a new shapshot is produced. In this fashion the computer generates a series of "stills" that can resemble the movement of cars in a very detailed fashion.

To heighten the drama you can introduce chance events: cars stalling, collisions, the driver who slows down to stare at the girl in the car next to him. Since chance events, by definition, are unpredictable, the computer decides when one will happen by its own chance mechanism; it computes a random number and compares it with a probability. For example, suppose the scenario calls for a new snapshot every half minute, and the average frequency of collisions is one every two hours. The likelihood of a collision occurring during any snapshot is one in two hundred and forty. A random number is generated, and if it is less than 1/240, a collision is recorded, and a special set of rules describe the kind of pile-up that will occur. If the random number is larger than 1/240, events proceed as normal. Models with chance events of this sort in them are called Monte-Carlo. (It's hard to escape from the gambling atmosphere where probabilities are concerned!)

In simulating law courts much the same procedure is followed, except that instead of a stream of cars we have a stream of "cases" and instead of moving along a highway,

they advance through several "stages"--arrest, initial appearance, preliminary hearing, grand jury indictment, arraignment, and court trial (and of course, everything moves much more slowly). Here the chance events could be the appearance of a case, its type, whether a judge was on vacation or ill, etc.

The nice thing about a simulation is that it is much like a laboratory, where experiments can be performed on complex systems at very little expense, and with none of the trouble of experimenting with the real situation. Variations can be played on procedures, rules, kinds of equipment, and the improvement (or disprovement) noted. A modern high-speed computer can go through hundreds of simulated traffic days in an hour, or try a dozen arrangements for an assembly line in a minute.

The Scientific Panel of the President's Commission on Law Enforcement conducted a simulation of a District of Columbia court where the primary interest was on the time required to process a case. A clear bottleneck showed up at the point of grand jury indictment. When the experiment was performed of adding a second, part-time grand jury, the time required for this stage was reduced from 35 days to less than 1 day.

Simulation is also good when you don't have a simple, well-balanced measure of what is good. You can experiment with your enterprise, look at the detailed results, and make a spot decision concerning the outcome that "looks best."

OPERATIONS RESEARCH AND SCIENCE

One of the marks of a young discipline is soul-searching concerning whether it is or is not a science; whether it has a solid and distinctive "method" or whether it has the traditional virtues of empirical verification, generality, and logical coherence. A glance at technical journals in the area of psychology in the first quarter of the century, or in sociology in the second, will reveal an astonishing number of papers devoted to agonizing on these questions. There was some of this in the early (fifteen years ago!) days of operations research, but it quickly damped out. The reason is that operations research has kept close to practical matters. Its development has resembled the growth of engineering or medicine, more than the development of academic science. The question has more often been "is it useful" than "is it true--or illuminating?"

This is as it should be. The steam engine came long before thermodynamics, and the telephone before information theory. Operations research is still very much pioneering in the areas of management technology and social engineering. In industry, operations research is steadily mounting the organization ladder, dealing with increasingly wider problems of executive interest. In government, the expansion of systems analysis from defense to problems of urban growth and crime can be expected to continue on to other social issues, even, as we have seen in the case of game theory, to issues of the structure of government itself. The pressures and allurements that these developments will exert on mathematics and the social sciences will be large. The theory will come, and long before the show is over. The outlines of this development are shadowy at the present time. But just as physical technology gave meaning and direction to the physical sciences, operations research can be expected to give orientation to the prescriptive sciences.

REFERENCES

1. Jensen, Arne, "Safety at Sea Problems," Proceedings of the Fourth International Conference on Operations Research, Boston, 1966.
2. Herman, R., and K. Gardelo, "Vehicular Traffic Flow," Scientific American, V. 209, No. 6, December 1963, pp. 35-43.
3. Ford, L. R., Jr., and D. R. Fulkerson, Flows in Networks Princeton University Press, Princeton, New Jersey, 1962.
4. Dantzig, George B., Linear Programming and Extension, The RAND Corporation, R-366, 1963.
5. Bellman, Richard, Dynamic Programming, Princeton University Press, Princeton, New Jersey, 1957.
6. Von Neumann, J., and O. Morgenstern, Theory of Games and Economic Behavior, Princeton University Press, Princeton, New Jersey, 1947.
7. Williams, J. D., The Compleat Strategyst, McGraw-Hill, New York, 1954.
8. Shapley, L. S., "A Value for n-person Games," contribution to The Theory of Games, H. W. Kuhn and A. W. Tucker, eds., Princeton, New Jersey, 1953.
9. Riker, W. H., and L. S. Shapley, Weighted Voting: A Mathematical Analysis for Instrumental Judgment, The RAND Corporation, P-3318, March 1966.